EXPLORATION OF QUALITY OF EXPERIENCE OF STEREOSCOPIC IMAGES: BINOCULAR DEPTH

Wei Chen, Fournier Jérôme, Marcus Barkowsky, Patrick Le Callet

EXPLORATION OF QUALITY OF EXPERIENCE OF STEREOSCOPIC IMAGES: BINOCULAR DEPTH

Wei Chen[1, 2], Jérôme Fournier[1], Marcus Barkowsky[2], Patrick Le Callet[2]
[1]Orange labs, France telecom R&D,
wei.chen, jerome.fournier@orange.com
[2]LUNAM Université, Université de Nantes, IRCCyN UMR CNRS 6597,
marcus.barkowsky, patrick.lecallet@univ-nantes.fr

ABSTRACT

Compared to traditional 2D image techniques, stereoscopic techniques provide additional information - the binocular depth, which strongly enhances the immersion. However, it may also cause visual comfort problems because it is still not a perfect representation of natural human vision but to some extent an illusion. How to fairly evaluate and understand the QoE (Quality of Experience) of stereoscopic images is still an open question. In this paper, we aim to explore the influence of binocular depth on the QoE of stereoscopic images by subjective quality assessment using methods that take into account the 3D concept. Firstly, quality indicators based on 3D concept including 2D image quality, depth rendering, depth quantity, visual comfort, naturalness and visual experience are defined. Synthetic scenes and natural scenes are designed and the camera parameters are calculated in order to create variation of perceived binocular depth range in terms of DoF (Depth of Focus). Subjective quality assessment based on the SAMVIQ method is used to evaluate the viewer's subjective opinions. The experiment results reveal how the variation of perceived binocular depth affects different quality aspects of 3D QoE. It leads to a proposal of modeling higher level concepts (depth rendering, naturalness and visual experience) of 3D QoE as a weighted sum of 2D image quality, depth rendering and visual experience in case of binocular depth variation.

1. INTRODUCTION

With the rapid development of optic and micro-conductor industry as well as the success of HDTV and digital cinema technique, there is almost no doubt that the next step of development of imaging system is targeting at 3D. However, advanced techniques such as holographic, which could represent a real 3D world to human vision system, are still far away from being mature enough so that stereoscopic technique, which can add one of the most important 3D information – the binocular depth, may be the most possible direct successor after 2D. The history of stereoscopic image system can be traced back to earlier 1838, the very first invent of stereoscopic imaging system [1]. In the past one hundred years, stereoscopic imaging techniques never stopped its development to provide better color, spatial and temporal resolution, yet some fundamental problems still can not be avoided and solved, e.g. the conflict of accommodation and convergence. QoE (Quality of Experience) is used to describe the human perceptual opinion for modern imaging system and subjective quality assessment is the traditional method to measure the QoE. However, 2D subjective quality assessment methods were proven to be insufficient to ,evaluate the 3D QoE[2], one of the main problems is that 2D quality indicators can not clearly highlight the added value such as stereoscopic depth and reveal problems such as visual discomfort in 3D. In [3], the author discussed the appropriate quality indicator for 3D assessment such as naturalness and viewing experience.

In this paper, we aim to focus on the exploration of how the most important added value - binocular depth variations affect the QoE of stereoscopic images. The paper is organized as follows: Section 2 defined the 3D QoE indicators for stereoscopic images which cover all the positive and negative quality aspects of stereoscopic images. Section 3 describes how the experiment contents (both synthetic scenes and natural scenes) were designed and captured in order to generate a variation of binocular depth. Section 4 focuses on the subjective quality assessment using SAMVIQ method[4] which revealed how binocular depth variation affect the different aspects of QoE on stereoscopic images. Section 5 models the higher level concept QoE indicators (depth rendering, naturalness and visual experience) as a weighted sum of basic elements (2D image quality, depth quantity and visual comfort). Concluding remarks are provided in Section 6.

2. QUALITY OF EXPERIENCE INDICATORS FOR STEREOSCOPIC IMAGE

The traditional concept to evaluate the quality of experience, i.e. the assessment of the overall visual quality, is not well adapted to assess the advantage and drawback of stereoscopic images, e.g. image quality is not sensitive to perceived depth and visual comfort problems. Seuntiëns et al [5] applied naturalness and viewing experience as QoE indicators to stereoscopic images in order to highlight the added value of stereoscopic images. Lambooij et al. in [6] stepped further to model these high evaluation concepts, i.e. the 3D QoE indicators, as a linear weighed sum of image quality and perceived depth. However, visual comfort problems were not evaluated and considered in their experiments. In our previous work [7], we evaluated the depth rendering, the visual comfort and the visual experience of stereoscopic images using the SAMVIQ method for synthetic stereoscopic images. The result revealed that increasing perceived depth would decrease the quality of the visual comfort, the visual experience and even the depth rendering itself.

In this paper, we aim to provide a more detailed subjective assessment to understand how binocular depth affects different quality aspects of stereoscopic images in terms of QoE.

The quality aspects of 3D QoE, i.e. the QoE indicators used in this paper are defined as below:

2D Image quality: the quality of rendering of texture, the level of visibility of visual artifacts and rendering details.

Depth quantity: the amount of the perceived depth using the combination of monocular and binocular depth cues.

Visual comfort: visual discomfort is related to multi-symptoms, e.g. eye strain, dry eyes, double vision. Variation of visual comfort can be perceived as the sensation of visual impairment as well as the sense of vision difficulties when moving the fixation point from one area of the image to another area (due to the decoupling of accommodation and convergence).

Depth rendering: the quality of the perceived depth, depending on the subject's preference on the basic criteria related to stretching or compression of the reality and the shape of the objects.

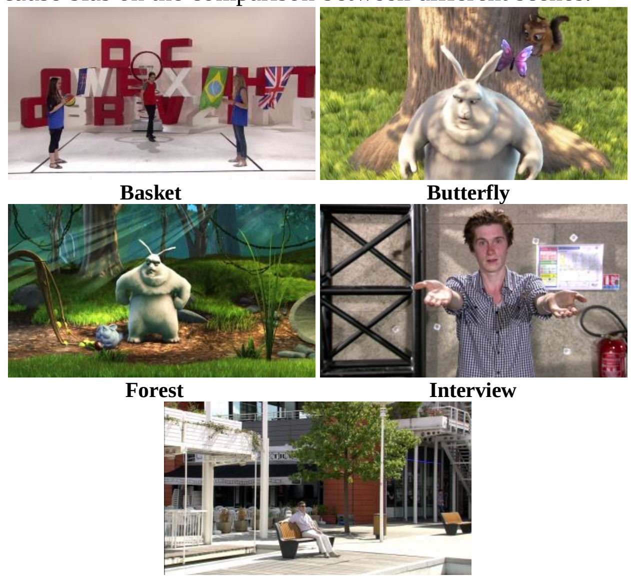
Naturalness: focuses on the evaluation of the natural appearance of images, i.e. whether the scene is more or less representative of reality.

Visual experience: the overall quality of experience of the images in terms of immersion and the overall perceived quality.

The observers in a subjective assessment are required to vote on all these possible visual opinions and it is believed that they merge into an overall QoE score.

3. STEREOSOCPIC IMAGE GENERATION AND CAPTURE

The final perceived depth of the stereoscopic images is depending on not only the shooting parameters but also on the visualization parameters[7, 8]. Moreover, the negative effect of stereoscopic imaging system, i.e. the visual discomfort, is also highly related to the final perceived depth range [9, 10]. In some previous research[11, 12], which the author captured the scenes of different depth range using a group of same camera baselines, which may result in different final perceived depth and which may cause bias on the comparison between different scenes.



Basket **Butterfly** **Forest** **Interview** **Bench**

Figure 1 Three natural scenes and two synthetic scenes

In this paper, the maximum final perceived depth range in the scene is represented as DoF (depth of focus) in the unit of diopters. All the camera parameters were calculated in order to represent the same final perceived depth range for each scene. Since DoF equal to 0.2 was proposed as the threshold of causing visual comfort problems [7, 10, 13, 14], for each scene, three DoF level, 0.1, 0.2 and 0.3 are captured and generated by adapting the shooting parameters (camera baseline) in order to represent the binocular depth variation. Both natural scenes and synthetic scenes were included. The natural scene capture used two professional 2D cameras (camera sensor 8.8x6.6 mm^2) and 3D rigs (mirror rig and side by side rig) in a toed-in setting. All the images were processed after capturing in order to avoid image asymmetry problems. The synthetic scene creation was based on the open animation project "big buck bunny"[15] and rendered by the Blender software (virtual camera sensor 32x16 mm^2).Three natural scenes and two synthetic scenes were used as shown in Figure 1 and all scene parameters are described in Table I.

Table I Scene parameters

Scene Name*	Near (m)	Far (m)	RoI* (m)	Conv* (m)
Basket(N)	5	10	7	5
Butterfly(S)	5.8	12	6.8	6.8
Forest(S)	5	23	7.5	5
Interview(N)	2.6	5	3	2.6
Bench(N)	<14	32	20	14

**N(Natural), S(Synthetic), RoI(Region of Interest), Conv(Convergence)*

Since the final visualization environment was selected as a 46 inch line interleaved display (Hyundai model S465D) and 4.5 times of display height viewing distance, Table II depicts the shooting parameters which were calculated to acquire the perceived binocular depth to guarantee DoF values as 0.1, 0.2 and 0.3 diopter in the final visualization.

Table II Shooting parameters

Scene Name	Focal (mm)	Camera baseline(mm)		
		DoF 0.1	DoF 0.2	DoF 0.3
Basket(N)	9	160	324	485
Butterfly(S)	70	118	236	353
Forest(S)	36	93	185	278
Interview(N)	22.5	35	65	105
Bench(N)	20	180	362	540

And the stereoscopic shape distortion factors[7], representing the shape distortion around the region of interest (a value of 1.0 indicates no shape distortion, less than 1 means compression in depth, larger than 1 means stretching in depth), are shown in Table III.

Table III Stereoscopic shape distortion

Scene Name	Stereoscopic shape distortion factor		
	DoF 0.1	DoF 0.2	DoF 0.3
Basket(N)	1	2.54	4.76
Butterfly(S)	0.69	1.38	2
Forest(S)	0.55	1.26	2.20
Interview(N)	0.5	1	1.78
Bench(N)	0.41	1.0	1.8

4. SUBJECTIVE QUALITY ASSESSMENT

The subjective quality assessment was designed as below:

1) **Test sessions:** the test consisted of six sessions corresponding to the six 3D QoE indicators defined in Section 2. In order to avoid interaction between the QoE indicators and in order to avoid accumulating visual discomfort, the whole test was separated into two parts which were conducted on two different days. The first part composed of three sessions: 2D image quality, depth rendering and visual comfort. The second part also consisted of three sessions: visual experiences, naturalness and depth quantity. For each session, there were 4 × 5 (DoF × scene) images presented to viewers for rating. The 20 stimuli were individually randomized for each perceptual scale.

2) **Equipment**: the subjective assessment was conducted in a test room, which is compliant with the recommendations for subjective evaluation of visual data issued by ITU-R BT.500 [16]. A 46 inch line-interleaved stereoscopic television with a native resolution of 1920x1080 pixels was used as the final visualization terminal. The viewing distance was fixed to 2.6 meter as 4.5 times of display height. The depth rendering ability of this display had been analyzed in [2] which showed an overall good depth rendering ability. A digital video system (DVS) which can output 1920x1080 HD signals was used to play the 3D content in a line interleaved mode.

3) Observers: 28 observers were recruited to participate in this test. All of them were non experts in the audiovisual and video domain. A vision test was performed on all testers to determine their visual performance and the potential impact on results. The test includes monocular visual acuity test, hyperopic trend, astigmatic trend, binocular distant vision acuity, dysphasia, fusion, stereoacuity and color vision. All observers had a normal or corrected to normal visual acuity and normal stereoacuity.

4) **Stimuli:** the image materials used in this experiment consisted of three natural scenes and two synthetic scenes are as shown in Figure 1. For each scene, there were four images representing the final perceived depth as DoF 0, 0.1, 0.2 and 0.3 diopters respectively as described in Section 2. The left view of the stereoscopic image representing 0.1 diopter DoF was used as a 2D image, also referred to as 0 diopter image. 4 × 5 (DoF × scene) images were presented in each test session.

5) **Procedure:** written instructions detailing the task what the observer had to perform and the attributes they were asked to rate were given to the subjects before the start of the test. These instructions were then reiterated by the experimenter as to ensure the observer understood the task. Subjective assessment methodology for Video Quality(SAMVIQ) [4] was used to evaluate each test condition on each perceptual scales. For each test session, five scenes, which have four stimuli in each scene, were evaluated by the subjects. For each scene, the subject could see all the four stimuli and rate their perceptual opinions. These stimuli were shown as buttons (A, B, C and D) and subjects could examine them respectively. The buttons were randomly reassigned to stimuli so that the subjects could not identify them. Each stimulus was shown with a fixed duration of 7s and subjects were asked to rate the evaluation criteria on a continuous scale labeled with the adjective terms [bad]-[poor]-[fair]-[good]-[excellent] according to the ITU-R BT.500. Specifically for the evaluation of depth quantity, a numerical scale from 0 to 100 was used. In this case, the subject was required to firstly identify the stimulus which had the largest depth sensation as 100 and they rated the other stimuli proportionally compared to this stimulus.

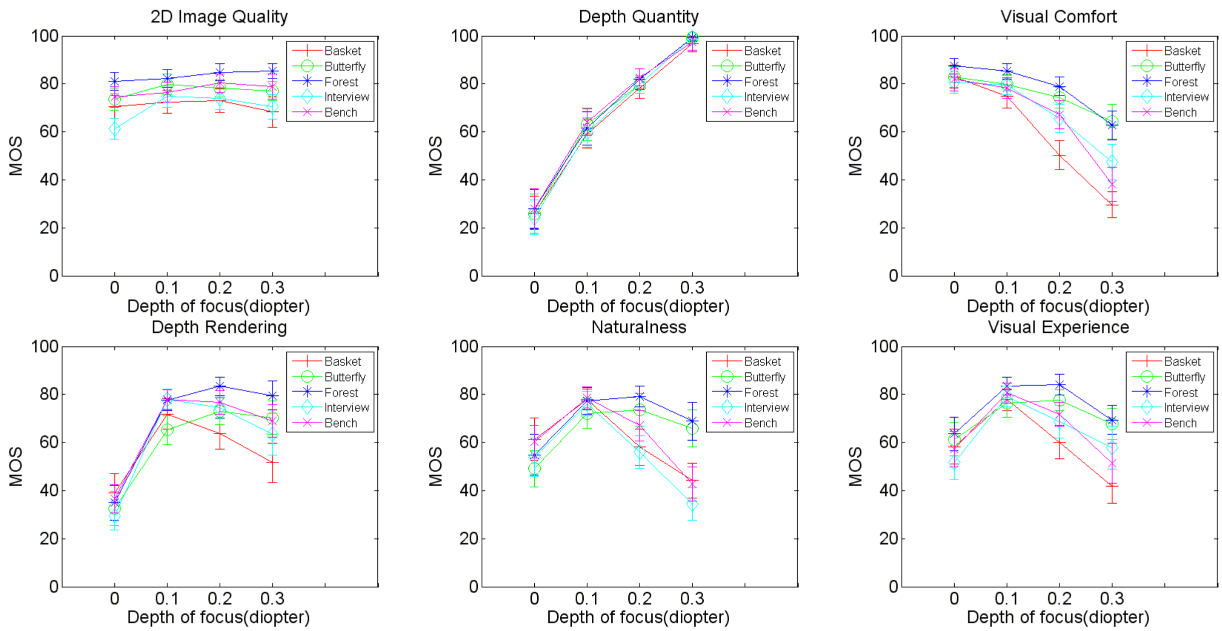


Figure 2 MOS (with their 95% confidence intervals) vs variation of DoF for different QoE indicators for different scenes (Baseket, Butterfly, Forest, Interview, and Bench as shown in Figure 1)

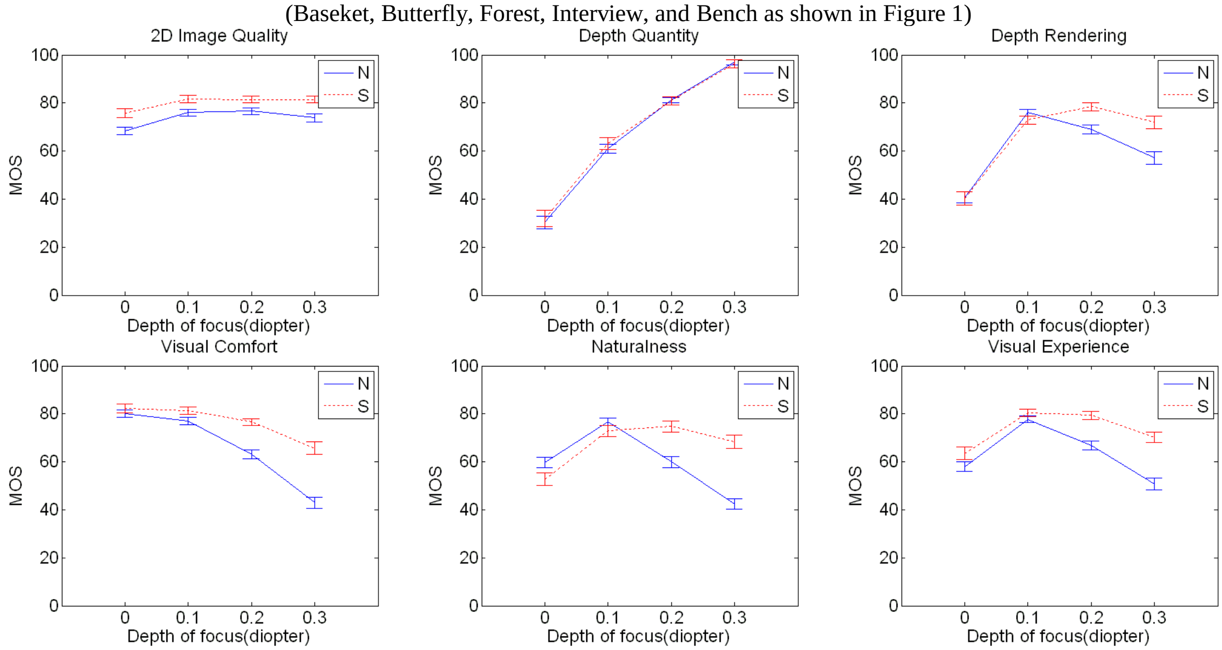


Figure 3 MOS (with their 95% confidence intervals) vs variation of DoF for different QoE indicators(Natural scene in solid line and Synthetic scene in dotted line)

Subjects were able to freely modify their scores within the 4 presented stimuli before continuing to the next scene.

6) **Results:** Figure 2 depicts the MOS (mean opinion score) with their 95% confidence intervals per quality indicator as a function of DoF (increasing along the x-axis) for each scene. An ANOVA analysis was performed with DoF as independent variable and MOS per quality indicator as dependent variable as well as two way interaction were included. The statistical analysis results show that image quality ($F = 0.96$, $p < 0.436$) is not affected by the variation of binocular depth. The result of depth quantity ($F = 1659$, $p < 0.001$) indicated that the subject can easily distinguish different perceived depth range. And with the increase of perceived depth, visual comfort ($F = 13.30$, $p < 0.001$), decreases significantly. Depth rendering ($F = 35.57$, $p < 0.001$), Naturalness ($F = 7.10$, $p < 0.004$) and Visual experience ($F = 9.496$, $p < 0.002$) all are similarly affected by the binocular depth variation. When increasing the perceived depth, in the beginning 3D shows advantages over 2D image, e.g. DoF

0 (as 2D) is rated as “poor” in depth rendering, and “fair” in naturalness and visual experience while in DoF 0.1 condition all of these indicators are scored between “good” and “excellent”. However, when the perceived depth is higher than a certain value (DoF 0.1 for Butterfly and Forest, DoF 0.2 for the other scenes), these advantages seem to be reduced. The feedback and discussion with the viewers confirmed that visual comfort should be the main concern which reduced the advantage of added depth as well as the shape distortion.

If we consider the shape distortion factor as shown in Table III, basket scene in DoF 0.1 and the other scene in DoF 0.2 should show advantages compared to the other perceived depth condition, especially in depth rendering. However, there are no significantly evidences shown in Figure 2 although in Basket (DoF 0.1), Butterfly (DoF 0.2), and Forest (DoF 0.2) the scales depth rendering, naturalness and visual experience are rated slightly better than the other conditions. This may be due to several reasons e.g. people are used to viewing 2D images and they are not sensitive to shape distortion in 3DTV especially in the case when the visual discomfort problem is essential.

Another finding from Figure 2 is that in terms of depth quantity and 2D image quality, all scenes behave similarly, yet in the other QoE indicators, the MOS for the natural scenes (Basket, Interview and Bench) drops faster than the synthetic scenes (Butterfly and Forest). Figure 3 depicts the MOS with their 95% confidence intervals per QoE indicator as a function of DoF between the natural scenes and synthetic scenes. For visual comfort, natural scenes decrease faster than synthetic scenes with the increase of DoF, e.g. in DoF 0.3, synthetic scenes still maintain “good” while natural scenes drop to some value between “fair” and “bad”. There are several possible explanations: firstly, human are used to viewing natural scene compared to synthetic scene; secondly, for natural shooting there exists some performance constrains such as optic focal length, thus blur effect cannot be avoided. For example, the background wall of the “interview” scene is strongly blurred and this blur may cause depth cue contradiction resulting in visual discomfort when people try to focus on the background. For synthetic scenes, all the contents were generated in a way that there appears no blur produced by the focal length and all depth layers are sharp. The same trends between the natural scenes and synthetic scenes are presented in depth rendering, naturalness and visual experience, which may be due to the interaction with visual comfort. This finding results in a recommendation for optimized perceived depth: For natural scenes DoF 0.1 should be targeted and for synthetic scenes the DOF may remain DoF 0.2[7].

5. MODELING THE 3D QOE

As explained in the previous section, 2D image quality is independent of depth variation yet depth quantity as well as visual comfort shows nearly linear relation with perceived binocular depth. Viewers can judge these three QoE indicators independently so that these entire three indicators may be categorized as the basic level of 3D QoE aspects. Furthermore, visual experience, naturalness and depth rendering may be defined as higher level of 3D QoE as people need to incorporate the basic level QoE concept in order to form the final perceptual opinion. This was revealed and discussed in Section 4. A 3D QoE model is proposed in Figure 4.

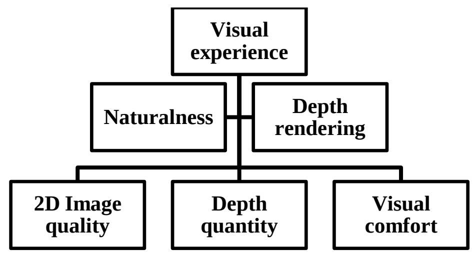


Figure 4 3D QoE model

Similar to [6], in order to explore the relationship between the higher level concept and the basic quality aspect in 3D QoE, we assume that the 3D QoE indicator (QoE) higher level quality indicators can be represented as a weighted sum of 2D image quality (IQ), depth quantity (D) and visual comfort (VC):

$$QoE = \alpha \cdot IQ + \beta \cdot D + \gamma \cdot VC$$

with α, β, γ representing the weights of 2D image quality, depth quantity and visual comfort respectively.

It should be noted that the current purpose of this paper is less relevant to modeling the 3D QoE by using physical parameters. Instead, the main purpose is to explore in which way high level 3D QoE is formed by basic level concepts. A simple standardized linear regression analysis was performed using the data shown in Figure 2 and the coefficients of each component for visual experience, naturalness and depth rendering are shown in the Table IV.

Table IV Standardized weighted coefficients

		IQ	D	VC	R square
Visual experience	Regression	0.028(a)	1.055	1,346	0.85
	Normalized(b)		0.44	0.56	
Naturalness	Regression	-0.005(a)	0.841	1.214	0.66
	Normalized(b)		0.41	0.59	
Depth rendering	Regression	0.015(a)	1.306	0.773	0.82
	Normalized(b)		0.63	0.37	

(a) not significant ;(b)normalized to a sum of one.

The linear fitting may be sufficient to explore the relationship between the higher level QoE concept and the basic level quality aspect as can be seen by the correlation coefficients(R square). All weighted coefficients of 2D image quality are close to 0 since it was not a significant factor for this dataset. This confirms the result that image quality is not affected by the variation of depth. Depth quantity and visual comfort are valid since they behave nearly linear with the variation of DoF.

The fitted coefficients show that depth quantity influences more on depth rendering (63%) than on visual experience (44%) and naturalness (41%) which also fits for its definition that required viewers to concentrate to the depth and space itself. Visual experience and naturalness scores are determined more by visual comfort (56% and 59% respectively) than by depth quantity. This also confirmed our optimal shooting rule [7] which defined that visual comfort is prior to perceived depth in order to guarantee a high overall QoE.

6. CONCLUSION

In this paper, we explored how binocular depth affects the quality of experience of stereoscopic images. Increasing the binocular depth does increase the perceived depth quantity as people can easily judge different perceived binocular depth levels. However, at the same time it decreased the visual comfort. 2D image quality is not affected by the variation of binocular depth. It was shown that the higher level quality indicators, depth rendering, naturalness and visual experience may be predicted by a weighted sum of image quality, depth quantity and visual comfort when only variation of binocular depth are considered. The coefficient of linear fitting showed that visual comfort is the dominant factor for visual experience (56%) and naturalness (59%), whereas depth rendering was determined by 37% from visual comfort and 63% from depth quantity.

7. ACKNOWLEGMENT

The authors would like to thank French National project - 3Dlive [17] for the support of the experiment content (natural scene) acquisition and post-production.

8. REFERENCES

[1] C. Wheatstone, "Contribution to the Physiology of Vision: Part the first, On some remarkable, and hitherto unobserved, Phenomena of Binocular Vision," *Philosophical Transactions 128*, 1838.

[2] W. Chen, J. Fournier, M. Barkowsky, and P. L. Callet, "New requirements of subjective video quality assessment methodlogies for 3DTV," in *Fifth International Workshop on Video Processing and Quality Metrics for Consumer Electronics - VPQM 2010*, Scottsdale, Arizona, U.S.A., 2010.

[3] P. Seuntiëns, "Visual experience of 3D TV," doctoral thesis, Eindhoven University of Technology, 2006.

[4] J. L. Blin, "New quality evaluation method suited to multimedia context SAMVIQ," in *The Second International Workshop on Video Processing and Quality Metrics for Consumer Electronic*, Phoenix,Arizona, 2006.

[5] P. J. Seuntiens, I. E. Heynderickx, W. A. Ijsselsteijn, P. M. J. van den Avoort, J. Berentsen, I. J. Dalm, M. T. Lambooij, and W. Oosting, "Viewing experience and naturalness of 3D images," in *Three-Dimensional TV, Video, and Display IV*, Boston, MA, USA, 2005, pp. 601605-7.

[6] M. Lambooij, W. Ijsselsteijn, D. G. Bouwhuis, and I. Heynderickx, "Evaluation of Stereoscopic Images: Beyond 2D Quality," *Broadcasting, IEEE Transactions on*, vol. 57, pp. 432-444, 2011.

[7] W. Chen, J. Fournier, M. Barkowsky, and P. L. Callet, "New stereoscopic video shooting rule based on stereoscopic distortion parameters and comfortable viewing zone," in *Stereoscopic Displays and Applications XXII*, San francisco, 2011.

[8] A. J. Woods, T. Docherty, and R. Koch, "Image distortions in stereoscopic video systems," in *Stereoscopic Displays and Applications IV*, San Jose, CA, USA, 1993, pp. 36-48.

[9] F. L. Kooi and A. Toet, "Visual comfort of binocular and 3D displays," *Displays*, vol. 25, pp. 99-108, 2004.

[10] S. Yano, S. Ide, T. Mitsuhashi, and H. Thwaites, "A study of visual fatigue and visual comfort for 3D HDTV/HDTV images," *Displays*, vol. 23, pp. 191-201, 2002.

[11] L. Goldmann, F. De Simone, and T. Ebrahimi, "A comprehensive database and subjective evaluation methodology for quality of experience in stereoscopic video," San Jose, California, USA, 2010, pp. 75260S-11.

[12] L. Goldmann, F. D. Simone, and T. Ebrahimi, "Impact of acquisition distortions on the quality of stereoscopic images," in *Fifth International Workshop on Video Processing and Quality Metrics for Consumer Electronics - VPQM 2010*, Scottsdale, Arizona, U.S.A., 2010.

[13] S. Yano, M. Emoto, and T. Mitsuhashi, "Two factors in visual fatigue caused by stereoscopic HDTV images," *Displays*, vol. 25, pp. 141-150, 2004.

[14] M. Lambooij, W. IJsselsteijn, and I. Heynderickx, "Stereoscopic displays and visual comfort: a review," Einhoven University of Technology Einhoven Univeristy of Technology Philips Research, SPIE Newsroom2 April 2007.

[15] Blender Foundation. (2008). *Big buck bunny*. Available: http://www.bigbuckbunny.org

[16] ITU, "Methodology for the subjective assessment of the quality of television pictures," in *Recommendation BT 500-11*, ed: International Telecommunication Union., 2002.

[17] *3DLIVE PROJECT*. Available online: http://3dlive-project.com